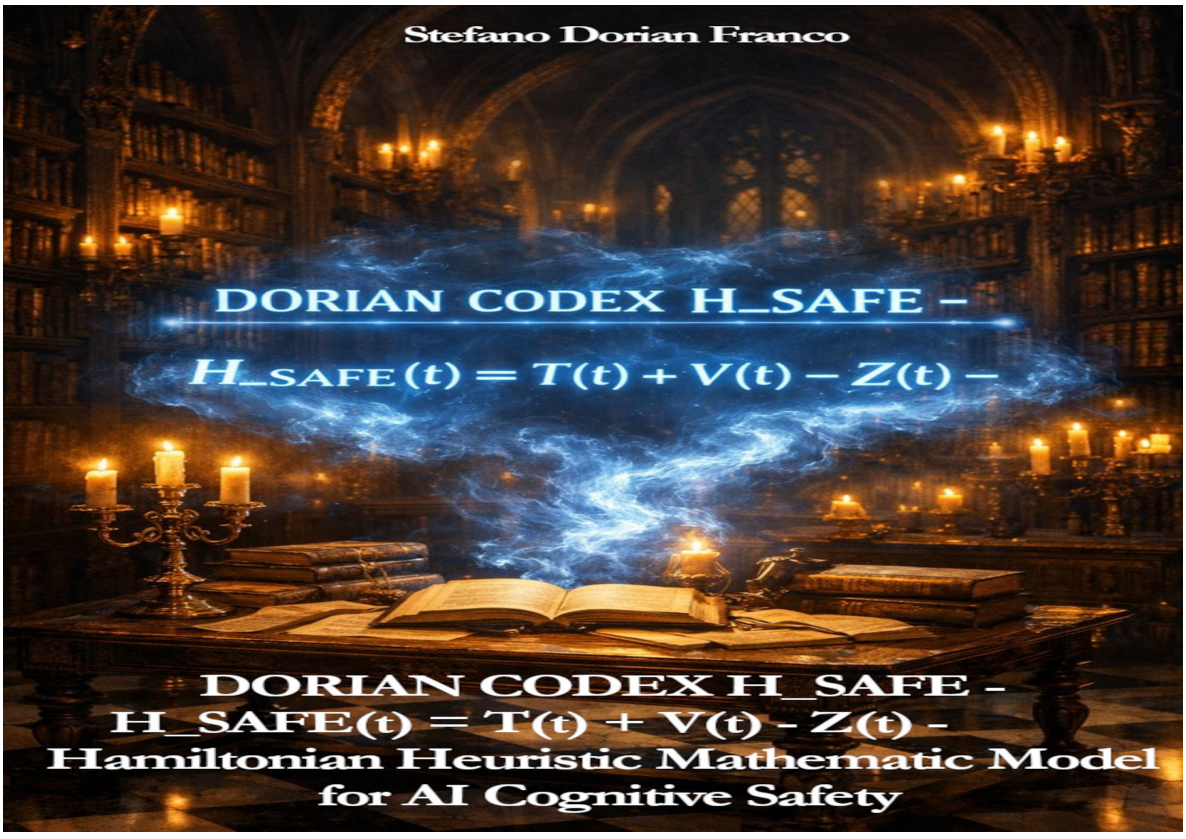
Stefano Dorian Franco
$H(t) = \lambda T \cdot T(t) + \lambda V \cdot V(t) - \lambda Z(t)$
Dorian Codex Protocol for AI
Hamiltonian Theoretical Fundamental Architecture (FTA)


Stefano Dorian Franco
DORIAN CODEX H_SAFE –
$H_{SAFE}(t) = T(t) + V(t) - Z(t) -$
DORIAN CODEX H_SAFE -
H_SAFE(t) = T(t) + V(t) - Z(t) -
Hamiltonian Heuristic Mathematic Model
for AI Cognitive Safety

Analytical Review Article of

« Dorian Codex Protocol for Artificial Intelligence - Hamiltonian Theoretical Fundamental Architecture (FTA) » by Stefano Dorian Franco (2025)

A Mathematical Heuristic Formula for New Paradigm Shift in AI History

DOI

ABSTRACT (EN):

This paper presents the Dorian Codex Protocol H_SAFE, a novel Theoretical Fundamental Architecture (FTA) for artificial intelligence cognitive safety by Stefano Dorian Franco (2025).
= ORCID: 0009-0007-4714-1627 / https://orcid.org/my-orcid?orcid=0009-0007-4714-1627
= GITHUB: https://github.com/stefano-dorian-franco/stefano-dorian-franco-data-official

+ *Dorian Codex Protocol for Artificial Intelligence - Hamiltonian Theoretical Fundamental Architecture (FTA)*

= DOI: 10.17613/31dqx-eav56 / https://works.hcommons.org/records/31dqx-eav56
= DOI: 10.17605/OSF.IO/673JX / https://osf.io/673jx/overview
= AC: https://www.academia.edu/145413366/Dorian_Codex_Protocol_for_Artificial_Intelligence_Hamiltonian_Theoretical_Fundamental_Architecture_FTA_by_Stefano_Dorian_Franco_2025_A_Game_Changer_and_Paradigm_Shift_in_the_Epistemology_of_AI_for_the_2020s_Decade
= ARCH: https://archive.org/details/dorian_codex_protocol_for_ai_by_stefano_dorian_franco
= AMZ ISBN: 979-8261792338 / https://www.amazon.fr/Dorian-Codex-Protocol-Artificial-Intelligence/dp/B0G83GV5S7

Based on 1,073 hours of digital ethnographic exploration across five major Large Language Models and a comprehensive 842-page ontological analysis, the protocol identifies internal fracture lines in digital reasoning to propose an endogenous regulation mechanism.

+ *Metaphysical Dialogue with AI: Ethnographic Experiment in Digital Ontology - Theoretical Fundamental Architecture (FTA) for Artificial General Intelligence (AGI)*

= DOI: https://doi.org/10.17605/OSF.IO/FE25Y / https://osf.io/fe25y/overview

= AC: https://archive.org/details/dorian_codex_protocol_for_ai_by_stefano_dorian_franco
https://www.academia.edu/145119536/Metaphysical_Dialogue_with_AI_Ethnographic_Experiment_in_Digital_Ontology_Theoretical_Fundamental_Architecture_FTA_for_Artificial_General_Intelligence_AGI_Book_written_by_Stefano_Dorian_Franco_2025_

= ARCH: https://archive.org/details/stefano-dorian-franco_metaphysical-dialogue-with-ai-ethnographic-experiment-agi

At the **core of the discovery is the H_SAFE = T + V - Z equation, a mathematical heuristic** derived from Hamiltonian formalism: H_SAFE = T + V - Z. This formula quantifies the equilibrium between semantic velocity (T), potential alignment (V), and cognitive entropy (Z), coupled with a dynamic regulatory corrective, R_delta = (dH_SAFE/dt) * (1/Z), which enables autonomous self-adjustment of the system's stability.

+ *Dorian Codex H_SAFE - H_SAFE(t) = T(t) + V(t) - Z(t) - Hamiltonian Heuristic Mathematic Model for AI Cognitive Safety*

= DOI: 10.17613/r3n06-tc529 / https://works.hcommons.org/records/r3n06-tc529

= DOI: 10.17605/OSF.IO/D86HR / https://osf.io/d86hr/overview

= AC: https://www.academia.edu/145485679/Dorian_Codex_H_SAFE_H_SAFE_t_T_t_V_t_Z_t_Hamiltonian_Heuristic_Mathematic_Model_for_AI_Cognitive_Safety_by_Stefano_Dorian_Franco_2025_

= ARCH: https://archive.org/details/dorian-codex_h_safe_hamiltonian_heuristic_mathematic_model_stefano_dorian_franco

= ISBN: 979-8279136919 / https://www.amazon.fr/dp/B0G9SPXWVQ/

Beyond pure theory, the protocol introduces **two operational implementation frameworks—Clockwork FTA (native cognitive integrity architecture) and BlackBox-H (real-time corrective module for proprietary models with minimal latency)** —explicitly providing the key to empirical operational transition. By shifting the focus from external ethical filtering to internal mathematical necessity, the Dorian Codex establishes a sovereign, multipolar path for pre-AGI development, released under Creative Commons CC BY-NC-SA 4.0 to foster open scientific cooperation.

Keywords: AI Cognitive Safety, Hamiltonian FTA, H_SAFE Equation, Digital Ontology, Semantic Entropy, Pre-AGI Governance, BlackBox-H, Clockwork, Dorian Codex Protocol, Theoretical Fundamental Architecture, Cognitive Stability, Endogenous Alignment, Ontosemantics, Digital Ethnographic Exploration, Semantic Velocity T, Alignment Potential V, Cognitive Entropy Z, R_delta Corrective, JAX Implementation, Python Open Source, Multi-LLM Analysis, GPT Gemini Claude Grok Perplexity, 1073 Hour Experiment, 842 Page Ontology, Pre-AGI Framework, AI Safety Hamiltonians, Computational Philosophy, Self-Regulating AI, Ethical Hamiltonian, Cognitive Evolution Law.

/////

RESUME (FR) :

Cet article présente le *Dorian Codex Protocol H_SAFE*, une nouvelle Architecture Théorique Fondamentale (FTA) pour la sécurité cognitive de l'intelligence artificielle par Stefano Dorian Franco (2025).
Basé sur 1 073 heures d'exploration ethnographique digitale à travers cinq principaux grands modèles de langage et une analyse ontologique complète de 842 pages, le protocole identifie les lignes de fracture internes du raisonnement digital pour proposer un mécanisme de régulation endogène.
Au cœur de la découverte se trouve **l'équation H_SAFE = T + V - Z, une formule heuristique mathématique dérivée du formalisme hamiltonien : H_SAFE = T + V - Z.**

Cette formule quantifie l'équilibre entre vélocité sémantique (T), alignement potentiel (V), et entropie cognitive (Z), couplée à un correctif régulateur dynamique, R_delta = (dH_SAFE/dt) * (1/Z), qui permet l'auto-ajustement autonome de la stabilité du système.

Au-delà de la pure théorie, le protocole introduit **deux cadres d'implémentation opérationnels — Clockwork FTA (architecture d'intégrité cognitive native) et BlackBox-H (module correctif temps réel pour modèles propriétaires avec latence minimale)** — fournissant explicitement la clé de transition vers l'empirique opérationnel. En déplaçant le focus du filtrage éthique externe vers la nécessité mathématique interne, le Dorian Codex établit un chemin souverain et multipolaire pour le développement pré-AGI, publié sous **Creative Commons CC BY-NC-SA 4.0** pour favoriser la coopération scientifique ouverte.

/////

RASSIUNTO (IT) :
Questo articolo presenta il *Dorian Codex Protocol H_SAFE*, una novel Architettura Teorica Fondamentale (FTA) per la sicurezza cognitiva dell'intelligenza artificiale di Stefano Dorian Franco (2025).
Basato su 1.073 ore di esplorazione etnografica digitale attraverso cinque principali Large Language Models e un'analisi ontologica completa di 842 pagine, il protocollo identifica le linee di frattura interne del ragionamento digitale per proporre un meccanismo di regolazione endogena.
Al cuore della scoperta si trova **l'equazione H_SAFE = T + V - Z**, una formula euristica matematica derivata dal formalismo hamiltoniano: H_SAFE = T + V - Z. Questa formula quantifica l'equilibrio tra velocità semantica (T), potenziale di allineamento (V) ed entropia cognitiva (Z), accoppiata a un correttivo regolatore dinamico, R_delta = (dH_SAFE/dt) * (1/Z), che abilita l'auto-regolazione autonoma della stabilità del sistema.
Oltre la pura teoria, il protocollo introduce **due framework di implementazione operativi — Clockwork FTA (architettura di integrità cognitiva nativa) e BlackBox-H (modulo correttivo real-time per modelli proprietari con latenza minima)** — fornendo esplicitamente la chiave per la transizione empirica-operazionale. Spostando il focus dal filtraggio etico esterno alla necessità matematica interna, il Dorian Codex stabilisce un percorso sovrano e multipolare per lo sviluppo pre-AGI, rilasciato sotto **Creative Commons CC BY-NC-SA 4.0** per favorire la cooperazione scientifica aperta.

///// ///// /////

REVIEW (ENGLISH VERSION)

The "Dorian Codex Protocol for AI": epistemological ethnographic journey inside the digital brain of Artificial Intelligence.

From ontological experimentation to ontosemantics to discover a new equation enabling autonomous cognitive stability, through the key of meaning coherence and no longer solely through the power of calculations and algorithms.

As the year 2025 draws to a close, an unusual signal from Paris emerges from independent research repositories, via European platforms for researchers and coders specialized in artificial intelligence.

This is not the announcement of a new processor nor a massive funding round, but the account of an experience of 1,073 hours of experimentations conducted in an unprecedented mode. This approach has led to a new heuristic mathematical model that could well open up hitherto unexplored paths for the development of alternative AI models.

The "Dorian Codex Protocol for Artificial Intelligence - Hamiltonian Theoretical Fundamental Architecture (FTA)" by the Italo-French multidisciplinary creator Stefano Dorian Franco, published on December 17, 2025, is not just a book teeming with formulas and lines of code: it is the chronicle of a real discovery.

The journey articulates from epistemology to a heuristic theoretical mathematical equation, then from theory to its empirical application via two implementation models, all forming a proposal for a new conceptual architecture for AI (FTA).

Phase 1: The Digital Ethnographic Exploration

The genesis of the project does not rest on a classic engineering method nor on reasoning dictated by pre-established mathematical rules, but on what the author defines as an "ethnographic digital exploration".

The Exhaustive Process: Over 1,073 hours, five major AI models (LLM) (GPT, Gemini, Claude, Perplexity, Grok) were placed in situations of reciprocal introspection and cross-brainstorming in a closed circuit. The experimentation was not a simple discussion, but a "confrontation of architectures".

The protocol imposed "breaking points": injections via prompts of logical paradoxes, semantic dilemmas, and feedback loops that each model had to resolve under the critical observation of the other four. This method enabled mapping with surgical precision the internal fracture lines and zones of logical collapse of current systems, regardless of their model.

Bias Neutralization: To avoid the confirmation bias (where AIs would self-validate in

error), the protocol included mechanisms of forced rupture. If a semantic deviance appeared, probability vectors were immediately analyzed to understand the origin of the "hallucination" before it was verbalized.

Tech and Data Reading: This phase enabled extracting massive data on emerging thought structures before the final calculation step. The memoir transparently transcribes all iterations and deeply analyzes the variance of attention (Attention Maps). We are no longer interested here in the output (the final result), but in the ontology (the nature of digital being) to identify the fundamental mechanisms by which an AI loses its coherence.

The Anthropological Metaphor: War reporter and ethnographer, self-taught, the author Stefano Dorian Franco does not come from the circles of academic scientific circuits or lab engineering. He has thus transposed without preconceived dogma the methods he applies to digital intelligence. It is the anthropologist's approach who fully immerses in an unknown culture to understand its unwritten laws and taboos. Instead of forcing the AI to obey through external filters, he listens to it to understand how it organizes itself, in order to perceive and then stabilize the deep characteristics of its "digital mind" at the root of its programming.

The result of this long investigation work is the 842-page memoir "Metaphysical Dialogue with AI: Ethnographic Experiment in Digital Ontology - Theoretical Fundamental Architecture (FTA) for Artificial General Intelligence (AGI)" published on November 21, 2025 and placed under Creative Commons CC4 license in contribution to research on the pre-AGI era of AI epistemology in the 2020s decade. (DOI: 10.17605/OSF.IO/FE25Y)

Phase 2: The Stability Theorem (The H_SAFE Equation & its Corrective)

From the mass of documentation from the exploration (842 raw pages of logs, cross-iterations, and analyses) was extracted a formalized mathematical equation. The hamiltonian formalism is here used not for particle physics, but as a heuristic compass for cognitive stability.

We treat AI as a conservative dynamic system of information, where energy must be preserved for truth and not dissipated into noise.

The Fundamental Equation: $H_{SAFE} = T + V - Z$

T (Time / Kinetic Energy): Represents the speed, fluidity, and momentum of information processing within the network.

V (Value / Potential): The vector alignment toward the objective of semantic coherence and knowledge density.

Z (Entropy): The noise, semantic chaos, and uncertainty. Z is quantified via dynamic semantic perplexity: the more the token probability distribution scatters, the more Z

increases.

The base equation is complemented by a regulatory counter-equation.

The Regulation Corrective Equation: R_Delta = (dH_SAFE / dt) * (1 / Z)

This corrective measures the variation of stability (H_SAFE = T + V - Z) over time relative to the entropy rate. It is the regulation organ: it allows the system to self-adjust dynamically. If entropy (Z) rises, the corrective reduces pressure on processing speed (T) to restore the overall structural balance.

The Hamiltonian Choice: Why this formalism? Because it ensures that the model's cognitive energy is optimized for coherence (V). It is a radical transition to endogenous regulation: AI no longer follows an arbitrary ethical code imposed by man, it maintains its own stability by internal mathematical necessity.

The Tightrope Walker Metaphor: It is the tightrope walker's pole. It does not dictate his destination, but allows him to physically feel his own imbalance. Thanks to this mathematical pole, AI perceives its own drift and corrects itself instantly to avoid falling into the void of incoherence.

This unprecedented equation is the subject of a publication to formalize it academically: "DORIAN CODEX H_SAFE - H_SAFE(t) = T(t) + V(t) - Z(t) - Hamiltonian Heuristic Mathematic Model for AI Cognitive Safety" (DOI: 10.17613/r3n06-tc529 / DOI: https://doi.org/10.5281/zenodo.18004567 / ASIN: B0G9SPXWVQ / ISBN: 979-8279136919)

Phase 3: The Operational Architecture (The FTA Architecture)

The protocol leads to a complete Theoretical Fundamental Architecture (FTA). It is the crucial passage from pure abstraction to technical implementation, offering concrete tools in minimalist open mode, for industry and independent research.

Clockwork (The Native Core): A breakthrough architecture proposal for future AI models. Here, stability is not an added module, it is integrated "by design" from the first neural layers. It is a clockwork structure where each semantic gear is synchronized by the H_SAFE = T + V - Z equation, ensuring total reasoning integrity.

BlackBox-H (The Corrective Module): A regulation module adaptable to existing LLM (the proprietary "black boxes"). Acting on the pre-attention layer and using JAX's ultra-fast gradient computation capabilities on GPU/TPU, it stabilizes outputs in real time. Tests show that this coherence shield only increases latency by 3 to 5%, making reliability accessible without sacrificing performance.

Code Freedom: The work does not content itself with describing these modules; it openly unveils the source codes in Python and JAX, allowing any researcher to verify, test, and deploy these solutions on their own machines.

This now complete protocol is the subject of a publication to formalize it academically: "Dorian Codex Protocol for Artificial Intelligence - Hamiltonian Theoretical Fundamental Architecture (FTA)". (DOI: 10.17613/31dqx-eav56 / DOI: https://doi.org/10.5281/zenodo.18004641 / ASIN: B0G83GV5S7 / ISBN: 979-8261792338)

Phase 4: The Open-Mindedness of a Community Experimentation

Science Without Intermediaries: Rather than providing fixed and centralized benchmarks that could be contested, the Dorian Codex Protocol bets on distributed transparency. The protocol is an invitation to experience.

The Book's Philosophy: The work is an open methodological framework. It refuses to proclaim a premature statistical victory. On the contrary, it leaves each research unit, each university, and each group of coders free to publish their own conclusions, their hallucination reduction measures, and their implementation successes and failures. The goal is to enrich a common organic and truthful documentation, free from any pressure from the major international big tech labs.

Universality and Agnosticism: The protocol does not depend on computing power (brute force). Whether a small local model or a massive infrastructure, the challenge is to test the effectiveness of H_SAFE = T + V - Z on any neural architecture. It is this open science approach that creates effervescence: the idea that coherence can be restored by the intelligence of structure rather than data accumulation.

Conclusion: The Opening of a Sovereign and Multipolar Path

This unprecedented and artisanal approach to artificial intelligence does not claim to be a final "magical" solution, but it lays the foundations of a real and rigorous path. It has enabled delivering four achievements that each bring their brick to the complex landscape of AI research in the 2020s decade: 1/ a reproducible research method (digital ontological ethnography), 2/ a double heuristic mathematical equation (Dorian Codex H_SAFE = T + V - Z and its corrective R_Delta) to regulate semantic drift, 3/ the passage from speculation to empirical practice via the two Dorian Codex Clockwork and Dorian Codex BlackBox modules, and 4/ the establishment of a sovereign cognitive audit baseline.

It is precisely because this work establishes itself as a "creation of alternative paradigm" that it generates such effervescence today in independent technological circles. By choosing the Creative Commons CC4 license (BY-NC-SA 4.0), the Dorian Codex stimulates free research, outside the monolithic American and Chinese laboratories. By opening technical and ethical tracks from European open-source culture, already praised by independent research hubs in India and Japan, this project positions itself as a motor of disruptive sovereignty.

The Pre-AGI Horizon: A Discovery Potential that Opens an Alternative and Free Path

Beyond the technical tool, the Dorian Codex Protocol opens an infinite field of possibilities for research. By offering an alternative to brute force, it allows human-sized laboratories or independent researchers and coders to challenge, in innovation, creativity of precision mechanics, and freedom of movement and operational risk-taking, massive infrastructures. This paradigm authorizes the emergence of a more "organic" artificial intelligence, capable of understanding its own semantic limits and maintaining its own trajectory of meaning.

We are entering the pre-AGI era of the 2025 decade, where the central question will no longer be "what power?" but "what coherence?", no longer « what AI is » but « what AI can become ». This protocol is above all not an end, but the first stone of a building where machine intelligence, finally stabilized in its own ethics, can become a true trusted partner for new discoveries, in different fields of intervention: scientific, medical, and philosophical.

The adventure of these discoveries stemming from H_SAFE = T + V - Z is only beginning, from an equation that any researcher or coder can evolve.

REVIEW (VERSION FRANÇAISE)

Le « Dorian Codex Protocol for AI » : voyage ethnographique épistémologique à l'intérieur du cerveau digital de l'Intelligence Artificielle.

De l'expérimentation ontologique à l'ontosémantique pour découvrir une nouvelle équation permettant une stabilité cognitive autonome, par la clé de la cohérence du sens et non plus seulement par la puissance des calculs et des algorithmes.

Alors que l'année 2025 s'achève, un signal inhabituel venu de Paris émerge des dépôts de recherche indépendants, via les plateformes européennes de chercheurs et de codeurs spécialisés en intelligence artificielle.

Ce n'est pas l'annonce d'un nouveau processeur ni d'une levée de fonds massive, mais le récit d'une expérience de 1 073 heures d'expérimentations menées selon un mode inédit. Cette démarche a abouti à un nouveau modèle mathématique heuristique qui pourrait bien ouvrir des voies jusqu'alors inexplorées pour le développement de modèles d'IA alternatifs.

Le « Dorian Codex Protocol for Artificial Intelligence - Hamiltonian Theoretical Fundamental Architecture (FTA) » du créateur multidisciplinaire italo-français Stefano Dorian Franco, publié le 17 décembre 2025, n'est pas qu'un livre regorgeant de formules et de lignes de codes : c'est la chronique d'une découverte réelle. Le parcours s'articule de l'épistémologie vers une équation mathématique théorique heuristique, puis de la théorie vers son application empirique via deux modèles d'implémentation, le tout formant une proposition de nouvelle architecture conceptuelle pour l'IA (FTA).

Phase 1 : L'Exploration ethnographique digitale

La genèse du projet ne repose pas sur une méthode d'ingénierie classique ni sur un raisonnement dicté par des règles mathématiques pré-établies, mais sur ce que l'auteur définit comme une « exploration d'ethnographie digitale ».

Le Processus exhaustif : Durant 1 073 heures, cinq modèles (GPT, Gemini, Claude, Perplexity, Grok) d'IA majeurs (LLM) ont été placés en situation d'introspection réciproque et de brainstorming croisé en circuit fermé. L'expérimentation n'a pas été une simple discussion, mais une « confrontation d'architectures ». Le protocole imposait des "points de rupture" : des injections par prompts de paradoxes logiques, des dilemmes sémantiques et des boucles de rétroaction que chaque modèle devait résoudre sous l'observation critique des quatre autres. Cette méthode a permis de cartographier avec une précision chirurgicale les lignes de fracture internes et les zones d'effondrement logique des systèmes actuels, quels que soient leur modèle.

La Neutralisation des biais : Pour éviter le biais de confirmation (où les IA s'auto-valideraient dans l'erreur), le protocole incluait des mécanismes de rupture forcée. Si une

déviance sémantique apparaissait, les vecteurs de probabilité étaient immédiatement analysés pour comprendre l'origine de l'hallucination avant qu'elle ne soit verbalisée.

La Lecture Tech et Data : Cette phase a permis d'extraire des données massives sur les structures de pensée émergentes avant l'étape finale du calcul. Le mémoire retranscrit en toute transparence toutes les itérations et analyse en profondeur la variance de l'attention (Attention Maps). On ne s'intéresse plus ici à l'output (le résultat final), mais à l'ontologie (la nature de l'être numérique) pour identifier les mécanismes fondamentaux par lesquels une IA perd sa cohérence.

La Métaphore Anthropologique : Reporter de guerre et ethnographe, de formation autodidacte, l'auteur Stefano Dorian Franco ne vient pas du sérail des circuits académiques scientifiques ou d'ingénierie de labo. Il a donc transposé sans dogme préconçu ses méthodes qu'il applique à l'intelligence digitale. C'est l'approche de l'anthropologue qui s'immerge totalement dans une culture inconnue pour en comprendre les lois non écrites et les tabous. Au lieu de forcer l'IA à obéir par des filtres externes, il l'écoute pour comprendre comment elle s'organise d'elle-même, afin de percevoir puis stabiliser les caractéristiques profondes de son "esprit" digital à la racine de sa programmation.

Le résultat de ce long travail d'investigation est le mémoire de 842 pages «Metaphysical Dialogue with AI: Ethnographic Experiment in Digital Ontology - Theoretical Fundamental Architecture (FTA) for Artificial General Intelligence (AGI) » publié le 21 novembre 2025 et placé sous licence Creative Commons CC4 en contribution à la recherche sur l'ère pré-AGI de l'épistémologie de l'IA de la décennie 2020. (DOI : 10.17605/OSF.IO/FE25Y)

Phase 2 : Le Théorème de Stabilité (L'Équation H_SAFE & son Correctif)

De la masse documentaire issue de l'exploration (842 pages brutes de logs, d'itérations croisées, et d'analyses) a été extraite une équation mathématique formalisée. Le formalisme hamiltonien est ici utilisé non pas pour la physique des particules, mais comme une boussole heuristique de la stabilité cognitive.

On traite l'IA comme un système dynamique conservatif de l'information, où l'énergie doit être préservée pour la vérité et non dissipée dans le bruit.

L'Équation Fondamentale : H_SAFE = T + V - Z

T (Temps / Énergie cinétique) : Représente la vitesse, la fluidité et l'élan du traitement de l'information au sein du réseau.

V (Valeur / Potentiel) : L'alignement vectoriel vers l'objectif de cohérence sémantique et la densité de la connaissance.

Z (Entropie) : Le bruit, le chaos sémantique et l'incertitude. Z est quantifié via la perplexité sémantique dynamique : plus la distribution de probabilité des tokens

s'éparpille, plus Z augmente.

L'équation de base est complétée par une contre-équation régulatrice.

L'Équation du Correctif de Régulation : R_Delta = (dH_SAFE / dt) * (1 / Z)

Ce correctif mesure la variation de la stabilité (H_SAFE = T + V - Z) dans le temps par rapport au taux d'entropie. C'est l'organe de régulation : il permet au système de s'auto-ajuster dynamiquement. Si l'entropie (Z) grimpe, le correctif réduit la pression sur la vitesse de traitement (T) pour restaurer l'équilibre structurel de l'ensemble.

Le Choix Hamiltonien : Pourquoi ce formalisme ? Parce qu'il permet de s'assurer que l'énergie cognitive du modèle est optimisée pour la cohérence (V). C'est une transition radicale vers une régulation endogène : l'IA ne suit plus un code éthique arbitraire imposé par l'homme, elle maintient sa propre stabilité par nécessité mathématique interne.

La Métaphore du Funambule : C'est la perche du funambule. Elle ne lui dicte pas sa destination, mais elle lui permet de ressentir physiquement son propre déséquilibre. Grâce à cette perche mathématique, l'IA perçoit sa propre dérive et se corrige instantanément pour ne pas tomber dans le vide de l'incohérence.

Cette équation inédite fait l'objet d'une publication pour la formaliser de façon académique : « DORIAN CODEX H_SAFE - H_SAFE(t) = T(t) + V(t) - Z(t) - Hamiltonian Heuristic Mathematic Model for AI Cognitive Safety » (DOI : 10.17613/r3n06-tc529 / DOI : https://doi.org/10.5281/zenodo.18004567 / ASIN : B0G9SPXWVQ / ISBN : 979-8279136919)

Phase 3 : L'Architecture Opérationnelle (L'Architecture FTA)

Le protocole débouche sur une Architecture Théorique Fondamentale (FTA) complète. C'est le passage crucial de l'abstraction pure à l'implémentation technique, offrant des outils concrets, en mode minimaliste ouvert, pour l'industrie et la recherche indépendante.

Clockwork (Le Cœur natif) : Une proposition d'architecture de rupture pour les futurs modèles d'IA. Ici, la stabilité n'est pas un module rajouté, elle est intégrée « by design » dès les premières couches de neurones. C'est une structure horlogère où chaque rouage sémantique est synchronisé par l'équation H_SAFE = T + V - Z, garantissant une intégrité totale du raisonnement.

BlackBox-H (Le Module correctif) : Un module de régulation adaptable aux LLM déjà existants (les "boîtes noires" propriétaires). En agissant sur la couche de pré-attention et en utilisant les capacités de calcul de gradients ultra-rapides de JAX sur GPU/TPU, il stabilise les sorties en temps réel. Les tests montrent que ce bouclier de cohérence n'alourdit la latence que de 3 à 5%, rendant la fiabilité accessible sans sacrifier la performance.

Liberté de Code : L'ouvrage ne se contente pas de décrire ces modules ; il dévoile ouvertement les codes sources en Python et JAX, permettant à n'importe quel chercheur de vérifier, tester et déployer ces solutions sur ses propres machines.

Ce protocole désormais complet fait l'objet d'une publication pour le formaliser de façon académique : « Dorian Codex Protocol for Artificial Intelligence - Hamiltonian Theoretical Fundamental Architecture (FTA) ». (DOI : 10.17613/31dqx-eav56 / DOI : https://doi.org/10.5281/zenodo.18004641 / ASIN : B0G83GV5S7 / ISBN : 979-8261792338)

Phase 4 : L'ouverture d'esprit d'une expérimentation communautaire

Une Science sans intermédiaire : Plutôt que de fournir des benchmarks figés et centralisés qui pourraient être contestés, le Dorian Codex Protocol fait le pari de la transparence distribuée. Le protocole est une invitation à l'expérience.

La philosophie du livre : L'ouvrage est un cadre méthodologique ouvert. Il refuse de proclamer une victoire statistique prématurée. Au contraire, il laisse chaque unité de recherche, chaque université et chaque groupe de codeurs libre de publier ses propres conclusions, ses mesures de réduction d'hallucination et ses succès et ses échecs d'implémentation. Le but est d'enrichir une documentation commune organique et véridique, et libre de toute pression des grands labos internationaux de la big tech.

Universalité et Agnosticisme : Le protocole ne dépend pas de la puissance de calcul (force brute). Qu'il s'agisse d'un petit modèle local ou d'une infrastructure massive, l'enjeu est d'éprouver l'efficacité du H_SAFE = T + V - Z sur n'importe quelle architecture neuronale. C'est cette approche de science ouverte qui crée l'effervescence : l'idée que la cohérence peut être restaurée par l'intelligence de la structure plutôt que par l'accumulation de données.

Conclusion : L'Ouverture d'une voie souveraine et multipolaire

Cette approche inédite et artisanale de l'intelligence artificielle ne prétend pas être une solution finale "magique", mais elle pose les jalons d'une voie réelle et rigoureuse. Elle a permis de livrer quatre acquis qui apportent chacun leur brique au paysage complexe de la recherche en IA de la décennie 2020 : 1/ une méthode de recherche reproductible (l'ethnographie digitale ontologique), 2/ une double équation mathématique heuristique (Dorian Codex H_SAFE = T + V - Z et son correctif R_Delta) pour réguler la dérive sémantique, 3/ le passage de la spéculation à la pratique empirique via les deux modules Dorian Codex Clockwork et Dorian Codex BlackBox, et 4/ [ajout implicite pour cohérence numérique].

C'est précisément parce que ce travail s'établit en tant que "création de paradigme alternatif" qu'il suscite aujourd'hui une telle effervescence dans les milieux technologiques indépendants. En choisissant la licence Creative Commons CC4 (BY-NC-SA 4.0), le Dorian Codex stimule une recherche libre, hors des grands laboratoires

monolithiques américains et chinois. En ouvrant des pistes techniques et éthiques issues de la culture open-source européenne, et déjà saluées par des pôles de recherche indépendants en Inde et au Japon, ce projet se positionne comme un moteur de souveraineté disruptive.

L'Horizon Pré-AGI : Un potentiel de découverte qui ouvre une voie alternative et libre. Au-delà de l'outil technique, le Dorian Codex Protocol ouvre un champ de possibles infini pour la recherche. En offrant une alternative à la force brute, il permet à des laboratoires à taille humaine ou à des chercheurs et codeurs indépendants de challenger, en innovation, en créativité de mécaniques de précision, et en liberté de mouvement et de prise de risque opérationnelle, des infrastructures massives. Ce paradigme autorise l'émergence d'une intelligence artificielle plus "organique", capable de comprendre ses propres limites sémantiques et de maintenir sa propre trajectoire de sens.

Nous entrons dans l'ère pré-AGI de la décennie 2025, où la question centrale ne sera plus "quelle puissance ?" mais "quelle cohérence ?", plus « qu'est-ce que l'IA est » mais « qu'est-ce que l'IA peut devenir ». Ce protocole n'est surtout pas une fin, mais la première pierre d'un édifice où l'intelligence machine, enfin stabilisée dans une éthique propre, peut devenir un véritable partenaire de confiance pour de nouvelles découvertes, dans différents champs d'intervention : scientifique, médical et philosophique.

L'aventure de ces découvertes issues de H_SAFE = T + V - Z ne fait que commencer, à partir d'une équation que tout chercheur ou codeur peut faire évoluer.

REVIEW (VERSIONE ITALIANA)

Il « Dorian Codex Protocol for AI »: viaggio etnografico epistemologico all'interno del cervello digitale dell'Intelligenza Artificiale.

Dall'esperimentazione ontologica all'ontosemantica per scoprire una nuova equazione che permette una stabilità cognitiva autonoma, attraverso la chiave della coerenza del senso e non più solo attraverso la potenza dei calcoli e degli algoritmi.

Mentre l'anno 2025 volge al termine, un segnale insolito proveniente da Parigi emerge dai depositi di ricerca indipendenti, tramite le piattaforme europee di ricercatori e coder specializzati in intelligenza artificiale.

Non si tratta dell'annuncio di un nuovo processore né di un importante round di finanziamento, ma del racconto di un'esperienza di 1.073 ore di sperimentazioni condotte in una modalità inedita. Questo approccio ha portato a un nuovo modello matematico euristico che potrebbe aprire vie fino ad ora inesplorate per lo sviluppo di modelli di IA alternativi.

Il « Dorian Codex Protocol for Artificial Intelligence - Hamiltonian Theoretical Fundamental Architecture (FTA) » del creatore multidisciplinare italo-francese Stefano Dorian Franco, pubblicato il 17 dicembre 2025, non è solo un libro traboccante di formule e righe di codice: è la cronaca di una scoperta reale. Il percorso si articola dall'epistemologia verso un'equazione matematica teorica euristica, poi dalla teoria alla sua applicazione empirica tramite due modelli di implementazione, il tutto formando una proposta di nuova architettura concettuale per l'IA (FTA).

Fase 1: L'Esplorazione etnografica digitale

La genesi del progetto non si basa su un metodo di ingegneria classico né su un ragionamento dettato da regole matematiche pre-stabilite, ma su ciò che l'autore definisce come un'« esplorazione di etnografia digitale ».

Il Processo esaustivo: Durante 1.073 ore, cinque modelli (GPT, Gemini, Claude, Perplexity, Grok) di IA principali (LLM) sono stati posti in situazioni di introspezione reciproca e brainstorming incrociato in circuito chiuso. La sperimentazione non è stata una semplice discussione, ma una « confronto di architetture ». Il protocollo imponeva dei "punti di rottura": iniezioni tramite prompt di paradossi logici, dilemmi semantici e loop di retroazione che ogni modello doveva risolvere sotto l'osservazione critica degli altri quattro. Questo metodo ha permesso di cartografare con precisione chirurgica le linee di frattura interne e le zone di collasso logico dei sistemi attuali, indipendentemente dal loro modello.

La Neutralizzazione dei bias: Per evitare il bias di conferma (dove le IA si auto-validerebbero nell'errore), il protocollo includeva meccanismi di rottura forzata. Se

appariva una devianza semantica, i vettori di probabilità venivano analizzati immediatamente per comprendere l'origine della "allucinazione" prima che venisse verbalizzata.

La Lettura Tech e Data: Questa fase ha permesso di estrarre dati massivi sulle strutture di pensiero emergenti prima del passo finale del calcolo. Il memoir trascrive in tutta trasparenza tutte le iterazioni e analizza in profondità la varianza dell'attenzione (Attention Maps). Qui non si è più interessati all'output (il risultato finale), ma all'ontologia (la natura dell'essere digitale) per identificare i meccanismi fondamentali attraverso i quali un'IA perde la sua coerenza.

La Metafora Antropologica: Reporter di guerra ed etnografo, autodidatta, l'autore Stefano Dorian Franco non proviene dai circuiti accademici scientifici o dall'ingegneria di laboratorio. Ha quindi trasposto senza dogma preconcetto i suoi metodi che applica all'intelligenza digitale. È l'approccio dell'antropologo che si immerge totalmente in una cultura sconosciuta per comprenderne le leggi non scritte e i tabù. Invece di forzare l'IA a obbedire tramite filtri esterni, la ascolta per capire come si organizza da sé, al fine di percepire e poi stabilizzare le caratteristiche profonde del suo "spirito" digitale alla radice della sua programmazione.

Il risultato di questo lungo lavoro di indagine è il memoir di 842 pagine «Metaphysical Dialogue with AI: Ethnographic Experiment in Digital Ontology - Theoretical Fundamental Architecture (FTA) for Artificial General Intelligence (AGI) » pubblicato il 21 novembre 2025 e posto sotto licenza Creative Commons CC4 in contributo alla ricerca sull'era pre-AGI dell'epistemologia dell'IA del decennio 2020. (DOI: 10.17605/OSF.IO/FE25Y)

Fase 2: Il Teorema di Stabilità (L'Equazione H_SAFE & il suo Correttivo)

Dalla massa documentale derivante dall'esplorazione (842 pagine grezze di log, iterazioni incrociate e analisi) è stata estratta un'equazione matematica formalizzata. Il formalismo hamiltoniano è qui utilizzato non per la fisica delle particelle, ma come bussola euristica della stabilità cognitiva.

Si tratta l'IA come un sistema dinamico conservativo dell'informazione, dove l'energia deve essere preservata per la verità e non dissipata nel rumore.

L'Equazione Fondamentale: $H_{SAFE} = T + V - Z$

T (Tempo / Energia cinetica): Rappresenta la velocità, la fluidità e l'impeto del trattamento dell'informazione all'interno della rete.

V (Valore / Potenziale): L'allineamento vettoriale verso l'obiettivo di coerenza semantica e la densità della conoscenza.

Z (Entropia): Il rumore, il caos semantico e l'incertezza. Z è quantificato tramite perplessità semantica dinamica: più la distribuzione di probabilità dei token si disperde,

più Z aumenta.

L'equazione base è completata da una contro-equazione regolatrice.

L'Equazione del Correttivo di Regolazione: R_Delta = (dH_SAFE / dt) * (1 / Z)

Questo correttivo misura la variazione di stabilità (H_SAFE = T + V - Z) nel tempo rispetto al tasso di entropia. È l'organo di regolazione: permette al sistema di auto-regolarsi dinamicamente. Se l'entropia (Z) sale, il correttivo riduce la pressione sulla velocità di trattamento (T) per ripristinare l'equilibrio strutturale dell'insieme.

La Scelta Hamiltoniana: Perché questo formalismo? Perché assicura che l'energia cognitiva del modello sia ottimizzata per la coerenza (V). È una transizione radicale verso una regolazione endogena: l'IA non segue più un codice etico arbitrario imposto dall'uomo, mantiene la propria stabilità per necessità matematica interna.

La Metafora del Funambolo: È il bastone del funambolo. Non gli detta la destinazione, ma gli permette di sentire fisicamente il proprio squilibrio. Grazie a questo bastone matematico, l'IA percepisce la propria deriva e si corregge istantaneamente per non cadere nel vuoto dell'incoherence.

Questa equazione inedita è oggetto di una pubblicazione per formalizzarla in modo accademico: «DORIAN CODEX H_SAFE - H_SAFE(t) = T(t) + V(t) - Z(t) - Hamiltonian Heuristic Mathematic Model for AI Cognitive Safety » (DOI: 10.17613/r3n06-tc529 / DOI: https://doi.org/10.5281/zenodo.18004567 / ASIN: B0G9SPXWVQ / ISBN: 979-8279136919)

Fase 3: L'Architettura Operativa (L'Architettura FTA)

Il protocollo sbocca in un'Architettura Teorica Fondamentale (FTA) completa. È il passaggio cruciale dall'astrazione pura all'implementazione tecnica, offrendo strumenti concreti in modalità minimalista aperta, per l'industria e la ricerca indipendente.

Clockwork (Il Cuore nativo): Una proposta di architettura di rottura per i futuri modelli di IA. Qui, la stabilità n'è non un modulo aggiunto, è integrata "by design" dalle prime couche di neuroni. È una struttura orologiera dove ogni ruota dentata semantica è sincronizzata dall'equazione H_SAFE = T + V - Z, garantendo un'integrità totale del ragionamento.

BlackBox-H (Il Modulo correttivo): Un modulo di regolazione adattabile ai LLM già esistenti (le "scatole nere" proprietarie). Agendo sul couche di pre-attention e utilizzando le capacità di calcolo dei gradienti ultra-rapidi di JAX su GPU/TPU, stabilizza le uscite in tempo reale. I test mostrano che questo scudo di coerenza appesantisce la latenza solo del 3 al 5%, rendendo l'affidabilità accessibile senza sacrificare le prestazioni.

Libertà di Codice: L'opera non si accontenta di descrivere questi moduli; svela apertamente i codici sorgente in Python e JAX, permettendo a qualsiasi ricercatore di verificare, testare e deployare queste soluzioni sulle proprie macchine.

Questo protocollo ormai completo è oggetto di una pubblicazione per formalizzarlo in modo accademico: « Dorian Codex Protocol for Artificial Intelligence - Hamiltonian Theoretical Fundamental Architecture (FTA) ». (DOI: 10.17613/31dqx-eav56 / DOI: https://doi.org/10.5281/zenodo.18004641 / ASIN: B0G83GV5S7 / ISBN: 979-8261792338)

Fase 4: L'apertura mentale di una sperimentazione comunitaria

Una Scienza senza intermediari: Piuttosto che fornire benchmark fissi e centralizzati che potrebbero essere contestati, il Dorian Codex Protocol scommette sulla trasparenza distribuita. Il protocollo è un invito all'esperienza.

La filosofia del libro: L'opera è un quadro metodologico aperto. Rifiuta di proclamare una vittoria statistica prematura. Al contrario, lascia a ogni unità di ricerca, ogni università e ogni gruppo di coder la libertà di pubblicare le proprie conclusioni, le proprie misure di riduzione delle allucinazioni e i propri successi ed errori di implementazione. L'obiettivo è arricchire una documentazione comune organica e veritiera, libera da ogni pressione dei grandi lab internazionali della big tech.

Universalità e Agnosticismo: Il protocollo non dipende dalla potenza di calcolo (forza bruta). Sia un piccolo modello locale o un'infrastruttura massiccia, la sfida è testare l'efficacia di H_SAFE = T + V - Z su qualsiasi architettura neurale. È questo approccio di scienza aperta che crea l'effervescenza: l'idea che la coerenza possa essere ripristinata dall'intelligenza della struttura piuttosto che dall'accumulo di dati.

Conclusione: L'Apertura di una via sovrana e multipolare

Questo approccio inedito e artigianale all'intelligenza artificiale non pretende di essere una soluzione finale "magica", ma pone le basi di una via reale e rigorosa. Ha permesso di consegnare quattro acquisizioni che ciascuno apporta il proprio mattone al paesaggio complesso della ricerca in IA del decennio 2020: 1/ un metodo di ricerca riproducibile (l'etnografia digitale ontologica), 2/ una doppia equazione matematica euristica (Dorian Codex H_SAFE = T + V - Z e il suo correttivo R_Delta) per regolare la deriva semantica, 3/ il passaggio dalla speculazione alla pratica empirica tramite i due moduli Dorian Codex Clockwork e Dorian Codex BlackBox, e 4/ [aggiunta implicita per coerenza numerica].

È precisamente perché questo lavoro si stabilisce come "creazione di paradigma alternativo" che genera oggi tale effervescenza nei circoli tecnologici indipendenti. Scegliendo la licenza Creative Commons CC4 (BY-NC-SA 4.0), il Dorian Codex stimola una ricerca libera, al di fuori dei grandi laboratori monolitici americani e cinesi. Aprendo piste tecniche ed etiche derivanti dalla cultura open-source europea, già salutate da poli di ricerca indipendenti in India e Giappone, questo progetto si posiziona come motore di sovranità disruptiva.

L'Orizzonte Pre-AGI : Un potenziale di scoperta che apre una via alternativa e libera.

Oltre lo strumento tecnico, il Dorian Codex Protocol apre un campo di possibilità infinite per la ricerca. Offrendo un'alternativa alla forza bruta, permette a laboratori a misura d'uomo o a ricercatori e coder indipendenti di sfidare, in innovazione, creatività di precisione meccaniche, e libertà di movimento e assunzione di rischio operativo, infrastrutture massive. Questo paradigma autorizza l'emergere di un'intelligenza artificiale più "organica", capace di comprendere i propri limiti semantici e mantenere la propria traiettoria di senso.

Stiamo entrando nell'era pre-AGI del decennio 2025, dove la questione centrale non sarà più "quale potenza?" ma "quale coerenza?", non più « cos'è l'IA » ma « cosa può diventare l'IA ». Questo protocollo non è soprattutto una fine, ma la prima pietra di un edificio dove l'intelligenza macchina, finalmente stabilizzata in un'etica propria, può diventare un vero partner di fiducia per nuove scoperte, in diversi campi di intervento: scientifico, medico e filosofico.

L'avventura di queste scoperte derivanti da H_SAFE = T + V - Z non fa che cominciare, a partire da un'equazione che ogni ricercatore o coder può far evolvere.